

# AUDIOVISUAL INFORMATION MANAGEMENT SYSTEM

## BACKGROUND OF THE INVENTION

5                   The present invention relates to a system for managing audiovisual information, and in particular to a system for audiovisual information browsing, filtering, searching, archiving, and personalization.

                  Video cassette recorders (VCRs) may record video programs in response to pressing a record button or may be programmed to record video programs based on the  
10                   time of day. However, the viewer must program the VCR based on information from a television guide to identify relevant programs to record. After recording, the viewer scans through the entire video tape to select relevant portions of the program for viewing using the functionality provided by the VCR, such as fast forward and fast reverse.

                  Unfortunately, the searching and viewing is based on a linear search, which may require  
15                   significant time to locate the desired portions of the program(s) and fast forward to the desired portion of the tape. In addition, it is time consuming to program the VCR in light of the television guide to record desired programs. Also, unless the viewer recognizes the programs from the television guide as desirable it is unlikely that the viewer will select such programs to be recorded.

20                   RePlayTV and TiVo have developed hard disk based systems that receive, record, and play television broadcasts in a manner similar to a VCR. The systems may be programmed with the viewer's viewing preferences. The systems use a telephone line interface to receive scheduling information similar to that available from a television guide.

5 Based upon the system programming and the scheduling information, the system automatically records programs that may be of potential interest to the viewer. Unfortunately, viewing the recorded programs occurs in a linear manner and may require substantial time. In addition, each system must be programmed for an individual's preference, likely in a different manner.

10 Freeman et al., U.S. Patent No. 5,861,881, disclose an interactive computer system where subscribers can receive individualized content.

With all the aforementioned systems, each individual viewer is required to program the device according to his particular viewing preferences. Unfortunately, each different type of device has different capabilities and limitations which limit the selections of the viewer. In addition, each device includes a different interface which the viewer may be unfamiliar with. Further, if the operator's manual is inadvertently misplaced it may be difficult for the viewer to efficiently program the device.

#### SUMMARY OF THE INVENTION

20 The present invention overcomes the aforementioned drawbacks of the prior art by providing at least one description scheme. For audio and/or video programs a program description scheme provides information regarding the associated program. For the user a user description scheme provides information regarding the user's preferences. For the system a system description scheme provides information regarding the system.

25 The description schemes are independent of one another. In the preferred embodiment the

5 system may use a combination of the description schemes to enhance its ability to search, filter, and browse audiovisual information in a personalized and effective manner.

The foregoing and other objectives, features and advantages of the invention will be more readily understood upon consideration of the following detailed description of the invention, taken in conjunction with the accompanying drawings.

## 10 BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is an exemplary embodiment of a program, a system, and a user, with associated description schemes, of an audiovisual system of the present invention.

FIG. 2 is an exemplary embodiment of the audiovisual system, including an analysis module, of FIG. 1.

15 FIG. 3 is an exemplary embodiment of the analysis module of FIG. 2.

FIG. 4 is an illustration of a thumbnail view (category) for the audiovisual system.

FIG. 5 is an illustration of a thumbnail view (channel) for the audiovisual system.

20 FIG. 6 is an illustration of a text view (channel) for the audiovisual system.

FIG. 7 is an illustration of a frame view for the audiovisual system.

FIG. 8 is an illustration of a shot view for the audiovisual system.

FIG. 9 is an illustration of a key frame view the audiovisual system.

FIG. 10 is an illustration of a highlight view for the audiovisual system.

25 FIG. 11 is an illustration of an event view for the audiovisual system.

5                   FIG. 12 is an illustration of a character/object view for the audiovisual system.

FIG. 13 is an alternative embodiment of a program description scheme including a syntactic structure description scheme, a semantic structure description scheme, a visualization description scheme, and a meta information description scheme.

10                  FIG. 14 is an exemplary embodiment of the visualization description scheme of FIG. 13.

FIG. 15 is an exemplary embodiment of the meta information description scheme of FIG. 13.

15                  FIG. 16 is an exemplary embodiment of a segment description scheme for the syntactic structure description scheme of FIG. 13.

FIG. 17 is an exemplary embodiment of a region description scheme for the syntactic structure description scheme of FIG. 13.

FIG. 18 is an exemplary embodiment of a segment/region relation description scheme for the syntactic structure description scheme of FIG. 13.

20                  FIG. 19 is an exemplary embodiment of an event description scheme for the semantic structure description scheme of FIG. 13.

FIG. 20 is an exemplary embodiment of an object description scheme for the semantic structure description scheme of FIG. 13.

25                  FIG. 21 is an exemplary embodiment of an event/object relation graph description scheme for the syntactic structure description scheme of FIG. 13.

5 DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

Many households today have many sources of audio and video information, such as multiple television sets, multiple VCR's, a home stereo, a home entertainment center, cable television, satellite television, internet broadcasts, world wide web, data services, specialized Internet services, portable radio devices, and a stereo in each of their  
10 vehicles. For each of these devices, a different interface is normally used to obtain, select, record, and play the video and/or audio content. For example, a VCR permits the selection of the recording times but the user has to correlate the television guide with the desired recording times. Another example is the user selecting a preferred set of preselected radio stations for his home stereo and also presumably selecting the same set of preselected  
15 stations for each of the user's vehicles. If another household member desires a different set of preselected stereo selections, the programming of each audio device would need to be reprogrammed at substantial inconvenience.

The present inventors came to the realization that users of visual information and listeners to audio information, such as for example radio, audio tapes,  
20 video tapes, movies, and news, desire to be entertained and informed in more than merely one uniform manner. In other words, the audiovisual information presented to a particular user should be in a format and include content suited to their particular viewing preferences. In addition, the format should be dependent on the content of the particular audiovisual information. The amount of information presented to a user or a listener  
25 should be limited to only the amount of detail desired by the particular user at the particular time. For example with the ever increasing demands on the user's time, the user may

5 desire to watch only 10 minutes of or merely the highlights of a basketball game. In  
addition, the present inventors came to the realization that the necessity of programming  
multiple audio and visual devices with their particular viewing preferences is a burdensome  
task, especially when presented with unfamiliar recording devices when traveling. When  
traveling, users desire to easily configure unfamiliar devices, such as audiovisual devices in  
10 a hotel room, with their viewing and listening preferences in a efficient manner.

The present inventors came to the further realization that a convenient  
technique of merely recording the desired audio and video information is not sufficient  
because the presentation of the information should be in a manner that is time efficient,  
especially in light of the limited time frequently available for the presentation of such  
15 information. In addition, the user should be able to access only that portion of all of the  
available information that the user is interested in, while skipping the remainder of the  
information.

A user is not capable of watching or otherwise listening to the vast potential  
amount of information available through all, or even a small portion of, the sources of  
20 audio and video information. In addition, with the increasing information potentially  
available, the user is not likely even aware of the potential content of information that he  
may be interested in. In light of the vast amount of audio, image, and video information,  
the present inventors came to the realization that a system that records and presents to the  
user audio and video information based upon the user's prior viewing and listening habits,  
25 preferences, and personal characteristics, generally referred to as user information, is  
desirable. In addition, the system may present such information based on the capabilities

5 of the system devices. This permits the system to record desirable information and to  
customize itself automatically to the user and/or listener. It is to be understood that user,  
viewer, and/or listener terms may be used interchangeability for any type of content. Also,  
the user information should be portable between and usable by different devices so that  
other devices may likewise be configured automatically to the particular user's preferences  
10 upon receiving the viewing information.

In light of the foregoing realizations and motivations, the present inventors  
analyzed a typical audio and video presentation environment to determine the significant  
portions of the typical audiovisual environment. First, referring to FIG. 1 the video, image,  
and/or audio information 10 is provided or otherwise made available to a user and/or a  
15 (device) system. Second, the video, image, and/or audio information is presented to the  
user from the system 12 (device), such as a television set or a radio. Third, the user  
interacts both with the system (device) 12 to view the information 10 in a desirable manner  
and has preferences to define which audio, image, and/or video information is obtained in  
accordance with the user information 14. After the proper identification of the different  
20 major aspects of an audiovisual system the present inventors then realized that information  
is needed to describe the informational content of each portion of the audiovisual system  
16.

With three portions of the audiovisual presentation system 16 identified, the  
functionality of each portion is identified together with its interrelationship to the other  
25 portions. To define the necessary interrelationships, a set of description schemes  
containing data describing each portion is defined. The description schemes include data

5 that is auxiliary to the programs 10, the system 12, and the user 14, to store a set of  
information, ranging from human readable text to encoded data, that can be used in  
enabling browsing, filtering, searching, archiving, and personalization. By providing a  
separate description scheme describing the program(s) 10, the user 14, and the system 12,  
the three portions (program, user, and system) may be combined together to provide an  
10 interactivity not previously achievable. In addition, different programs 10, different users  
14, and different systems 12 may be combined together in any combination, while still  
maintaining full compatibility and functionality. It is to be understood that the description  
scheme may contain the data itself or include links to the data, as desired.

A program description scheme 18 related to the video, still image, and/or  
15 audio information 10 preferably includes two sets of information, namely, program views  
and program profiles. The program views define logical structures of the frames of a video  
that define how the video frames are potentially to be viewed suitable for efficient  
browsing. For example the program views may contain a set of fields that contain data for  
the identification of key frames, segment definitions between shots, highlight definitions,  
20 video summary definitions, different lengths of highlights, thumbnail set of frames,  
individual shots or scenes, representative frame of the video, grouping of different events,  
and a close-up view. The program view descriptions may contain thumbnail, slide, key  
frame, highlights, and close-up views so that users can filter and search not only at the  
program level but also within a particular program. The description scheme also enables  
25 users to access information in varying detail amounts by supporting, for example, a key  
frame view as a part of a program view providing multiple levels of summary ranging from



5 .coarse to fine. The program profiles define distinctive characteristics of the content of the  
program, such as actors, stars, rating, director, release date, time stamps, keyword  
identification, trigger profile, still profile, event profile, character profile, object profile,  
color profile, texture profile, shape profile, motion profile, and categories. The program  
profiles are especially suitable to facilitate filtering and searching of the audio and video  
10 information. The description scheme enables users to have the provision of discovering  
interesting programs that they may be unaware of by providing a user description scheme.  
The user description scheme provides information to a software agent that in turn performs  
a search and filtering on behalf of the user by possibly using the system description scheme  
and the program description scheme information. It is to be understood that in one of the  
15 embodiments of the invention merely the program description scheme is included.

Program views contained in the program description scheme are a feature  
that supports a functionality such as close-up view. In the close-up view, a certain image  
object, e.g., a famous basketball player such as Michael Jordan, can be viewed up close by  
playing back a close-up sequence that is separate from the original program. An alternative  
20 view can be incorporated in a straightforward manner. Character profile on the other hand  
may contain spatio-temporal position and size of a rectangular region around the character  
of interest. This region can be enlarged by the presentation engine, or the presentation  
engine may darken outside the region to focus the user's attention to the characters  
spanning a certain number of frames. Information within the program description scheme  
25 may contain data about the initial size or location of the region, movement of the region  
from one frame to another, and duration and terms of the number of frames featuring the

5 region. The character profile also provides provision for including text annotation and audio annotation about the character as well as web page information, and any other suitable information. Such character profiles may include the audio annotation which is separate from and in addition to the associated audio track of the video.

The program description scheme may likewise contain similar information  
10 regarding audio (such as radio broadcasts) and images (such as analog or digital photographs or a frame of a video).

The user description scheme 20 preferably includes the user's personal preferences, and information regarding the user's viewing history such as for example browsing history, filtering history, searching history, and device setting history. The user's  
15 personal preferences includes information regarding particular programs and categorizations of programs that the user prefers to view. The user description scheme may also include personal information about the particular user, such as demographic and geographic information, e.g. zip code and age. The explicit definition of the particular programs or attributes related thereto permits the system 16 to select those programs from  
20 the information contained within the available program description schemes 18 that may be of interest to the user. Frequently, the user does not desire to learn to program the device nor desire to explicitly program the device. In addition, the user description scheme 20 may not be sufficiently robust to include explicit definitions describing all desirable programs for a particular user. In such a case, the capability of the user description scheme  
25 20 to adapt to the viewing habits of the user to accommodate different viewing characteristics not explicitly provided for or otherwise difficult to describe is useful. In

5 such a case, the user description scheme 20 may be augmented or any technique can be  
used to compare the information contained in the user description scheme 20 to the  
available information contained in the program description scheme 18 to make selections.  
The user description scheme provides a technique for holding user preferences ranging  
from program categories to program views, as well as usage history. User description  
10 scheme information is persistent but can be updated by the user or by an intelligent  
software agent on behalf of the user at any arbitrary time. It may also be disabled by the  
user, at any time, if the user decides to do so. In addition, the user description scheme is  
modular and portable so that users can carry or port it from one device to another, such as  
with a handheld electronic device or smart card or transported over a network connecting  
15 multiple devices. When user description scheme is standardized among different  
manufacturers or products, user preferences become portable. For example, a user can  
personalize the television receiver in a hotel room permitting users to access information  
they prefer at any time and anywhere. In a sense, the user description scheme is persistent  
and timeless based. In addition, selected information within the program description  
20 scheme may be encrypted since at least part of the information may be deemed to be  
private (e.g., demographics). A user description scheme may be associated with an  
audiovisual program broadcast and compared with a particular user's description scheme of  
the receiver to readily determine whether or not the program's intended audience profile  
matches that of the user. It is to be understood that in one of the embodiments of the  
25 invention merely the user description scheme is included.

5                   The system description scheme 22 preferably manages the individual  
programs and other data. The management may include maintaining lists of programs,  
categories, channels, users, videos, audio, and images. The management may include the  
capabilities of a device for providing the audio, video, and/or images. Such capabilities  
may include, for example, screen size, stereo, AC3, DTS, color, black/white, etc. The  
10   management may also include relationships between any one or more of the user, the  
audio, and the images in relation to one or more of a program description scheme(s) and a  
user description scheme(s). In a similar manner the management may include relationships  
between one or more of the program description scheme(s) and user description scheme(s).  
It is to be understood that in one of the embodiments of the invention merely the system  
15   description scheme is included.

                  The descriptors of the program description scheme and the user description  
scheme should overlap, at least partially, so that potential desirability of the program can be  
determined by comparing descriptors representative of the same information. For example,  
the program and user description scheme may include the same set of categories and actors.  
20   The program description scheme has no knowledge of the user description scheme, and  
vice versa, so that each description scheme is not dependant on the other for its existence.  
It is not necessary for the description schemes to be fully populated. It is also beneficial  
not to include the program description scheme with the user description scheme because  
there will likely be thousands of programs with associated description schemes which if  
25   combined with the user description scheme would result in a unnecessarily large user  
description scheme. It is desirable to maintain the user description scheme small so that it

5 is more readily portable. Accordingly, a system including only the program description scheme and the user description scheme would be beneficial.

The user description scheme and the system description scheme should include at least partially overlapping fields. With overlapping fields the system can capture the desired information, which would otherwise not be recognized as desirable. The  
10 system description scheme preferably includes a list of users and available programs.

Based on the master list of available programs, and associated program description scheme, the system can match the desired programs. It is also beneficial not to include the system description scheme with the user description scheme because there will likely be thousands of programs stored in the system description schemes which if combined with the user  
15 description scheme would result in a unnecessarily large user description scheme. It is desirable to maintain the user description scheme small so that it is more readily portable.

For example, the user description scheme may include radio station preselected frequencies and/or types of stations, while the system description scheme includes the available stations for radio stations in particular cities. When traveling to a different city the user description  
20 scheme together with the system description scheme will permit reprogramming the radio stations. Accordingly, a system including only the system description scheme and the user description scheme would be beneficial.

The program description scheme and the system description scheme should include at least partially overlapping fields. With the overlapping fields, the system  
25 description scheme will be capable of storing the information contained within the program description scheme, so that the information is properly indexed. With proper indexing, the

5 system is capable of matching such information with the user information, if available, for  
obtaining and recording suitable programs. If the program description scheme and the  
system description scheme were not overlapping then no information would be extracted  
from the programs and stored. System capabilities specified within the system description  
scheme of a particular viewing system can be correlated with a program description  
10 scheme to determine the views that can be supported by the viewing system. For instance,  
if the viewing device is not capable of playing back video, its system description scheme  
may describe its viewing capabilities as limited to keyframe view and slide view only.  
Program description scheme of a particular program and system description scheme of the  
viewing system are utilized to present the appropriate views to the viewing system. Thus, a  
15 server of programs serves the appropriate views according to a particular viewing system's  
capabilities, which may be communicated over a network or communication channel  
connecting the server with user's viewing device. It is preferred to maintain the program  
description scheme separate from the system description scheme because the content  
providers repackage the content and description schemes in different styles, times, and  
20 formats. Preferably, the program description scheme is associated with the program, even  
if displayed at a different time. Accordingly, a system including only the system  
description scheme and the program description scheme would be beneficial.

By preferably maintaining the independence of each of the three description  
schemes while having fields that correlate the same information, the programs 10, the users  
25 14, and the system 12 may be interchanged with one another while maintaining the  
functionality of the entire system 16. Referring to FIG. 2, the audio, visual, or audiovisual

5 program 38, is received by the system 16. The program 38 may originate at any suitable source, such as for example broadcast television, cable television, satellite television, digital television, Internet broadcasts, world wide web, digital video discs, still images, video cameras, laser discs, magnetic media, computer hard drive, video tape, audio tape, data services, radio broadcasts, and microwave communications. The program description  
10 stream may originate from any suitable source, such as for example PSIP/DVB-SI information in digital television broadcasts, specialized digital television data services, specialized Internet services, world wide web, data files, data over the telephone, and memory, such as computer memory. The program, user, and/or system description scheme may be transported over a network (communication channel). For example, the system  
15 description scheme may be transported to the source to provide the source with views or other capabilities that the device is capable of using. In response, the source provides the device with image, audio, and/or video content customized or otherwise suitable for the particular device. The system 16 may include any device(s) suitable to receive any one or more of such programs 38. An audiovisual program analysis module 42 performs an  
20 analysis of the received programs 38 to extract and provide program related information (descriptors) to the description scheme (DS) generation module 44. The program related information may be extracted from the data stream including the program 38 or obtained from any other source, such as for example data transferred over a telephone line, data already transferred to the system 16 in the past, or data from an associated file. The  
25 program related information preferably includes data defining both the program views and the program profiles available for the particular program 38. The analysis module 42

5 performs an analysis of the programs 38 using information obtained from (i) automatic audio-video analysis methods on the basis of low-level features that are extracted from the program(s), (ii) event detection techniques, (iii) data that is available (or extractable) from data sources or electronic program guides (EPGs, DVB-SI, and PSIP), and (iv) user information obtained from the user description scheme 20 to provide data defining the  
10 program description scheme.

The selection of a particular program analysis technique depends on the amount of readily available data and the user preferences. For example, if a user prefers to watch a 5 minute video highlight of a particular program, such as a basketball game, the analysis module 42 may invoke a knowledge based system 90 (FIG. 3) to determine the  
15 highlights that form the best 5 minute summary. The knowledge based system 90 may invoke a commercial filter 92 to remove commercials and a slow motion detector 54 to assist in creating the video summary. The analysis module 42 may also invoke other modules to bring information together (e.g., textual information) to author particular program views. For example, if the program 38 is a home video where there is no further  
20 information available then the analysis module 42 may create a key-frame summary by identifying key-frames of a multi-level summary and passing the information to be used to generate the program views, and in particular a key frame view, to the description scheme. Referring also to FIG. 3, the analysis module 42 may also include other sub-modules, such as for example, a de-mux/decoder 60, a data and service content analyzer 62, a text  
25 processing and text summary generator 64, a close caption analyzer 66, a title frame



5 generator 68, an analysis manager 70, an audiovisual analysis and feature extractor 72, an event detector 74, a key-frame summarizer 76, and a highlight summarizer 78.

The generation module 44 receives the system information 46 for the system description scheme. The system information 46 preferably includes data for the system description scheme 22 generated by the generation module 44. The generation module 44  
10 also receives user information 48 including data for the user description scheme. The user information 48 preferably includes data for the user description scheme generated within the generation module 44. The user input 48 may include, for example, meta information to be included in the program and system description scheme. The user description scheme (or corresponding information) is provided to the analysis module 42 for selective analysis  
15 of the program(s) 38. For example, the user description scheme may be suitable for triggering the highlight generation functionality for a particular program and thus generating the preferred views and storing associated data in the program description scheme. The generation module 44 and the analysis module 42 provide data to a data storage unit 50. The storage unit 50 may be any storage device, such as memory or  
20 magnetic media.

A search, filtering, and browsing (SFB) module 52 implements the description scheme technique by parsing and extracting information contained within the description scheme. The SFB module 52 may perform filtering, searching, and browsing of the programs 38, on the basis of the information contained in the description schemes.  
25 An intelligent software agent is preferably included within the SFB module 52 that gathers and provides user specific information to the generation module 44 to be used in authoring

5 and updating the user description scheme (through the generation module 44). In this manner, desirable content may be provided to the user through a display 80. The selections of the desired program(s) to be retrieved, stored, and/or viewed may be programmed, at least in part, through a graphical user interface 82. The graphical user interface may also include or be connected to a presentation engine for presenting the information to the user  
10 through the graphical user interface.

The intelligent management and consumption of audiovisual information using the multi-part description stream device provides a next-generation device suitable for the modern era of information overload. The device responds to changing lifestyles of individuals and families, and allows everyone to obtain the information they desire anytime  
15 and anywhere they want.

An example of the use of the device may be as follows. A user comes home from work late Friday evening being happy the work week is finally over. The user desires to catch up with the events of the world and then watch ABC's 20/20 show later that evening. It is now 9 PM and the 20/20 show will start in an hour at 10 PM. The user is  
20 interested in the sporting events of the week, and all the news about the Microsoft case with the Department of Justice. The user description scheme may include a profile indicating a desire that the particular user wants to obtain all available information regarding the Microsoft trial and selected sporting events for particular teams. In addition, the system description scheme and program description scheme provide information  
25 regarding the content of the available information that may selectively be obtained and recorded. The system, in an autonomous manner, periodically obtains and records the

5 audiovisual information that may be of interest to the user during the past week based on  
the three description schemes. The device most likely has recorded more than one hour of  
audiovisual information so the information needs to be condensed in some manner. The  
user starts interacting with the system with a pointer or voice commands to indicate a  
desire to view recorded sporting programs. On the display, the user is presented with a list  
10 of recorded sporting events including Basketball and Soccer. Apparently the user's  
favorite Football team did not play that week because it was not recorded. The user is  
interested in basketball games and indicates a desire to view games. A set of title frames is  
presented on the display that captures an important moment of each game. The user selects  
the Chicago Bulls game and indicates a desire to view a 5 minute highlight of the game.  
15 The system automatically generates highlights. The highlights may be generated by audio  
or video analysis, or the program description scheme includes data indicating the frames  
that are presented for a 5 minute highlight. The system may have also recorded web-based  
textual information regarding the particular Chicago-Bulls game which may be selected by  
the user for viewing. If desired, the summarized information may be recorded onto a  
20 storage device, such as a DVD with a label. The stored information may also include an  
index code so that it can be located at a later time. After viewing the sporting events the  
user may decide to read the news about the Microsoft trial. It is now 9:50 PM and the user  
is done viewing the news. In fact, the user has selected to delete all the recorded news  
items after viewing them. The user then remembers to do one last thing before 10 PM in  
25 the evening. The next day, the user desires to watch the VHS tape that he received from  
his brother that day, containing footage about his brother's new baby girl and his vacation

5 to Peru last summer. The user wants to watch the whole 2-hour tape but he is anxious to see what the baby looks like and also the new stadium built in Lima, which was not there last time he visited Peru. The user plans to take a quick look at a visual summary of the tape, browse, and perhaps watch a few segments for a couple of minutes, before the user takes his daughter to her piano lesson at 10 AM the next morning. The user plugs in the  
10 tape into his VCR, that is connected to the system, and invokes the summarization functionality of the system to scan the tape and prepare a summary. The user can then view the summary the next morning to quickly discover the baby's looks, and playback segments between the key-frames of the summary to catch a glimpse of the crying baby. The system may also record the tape content onto the system hard drive (or storage device)  
15 so the video summary can be viewed quickly. It is now 10:10 PM, and it seems that the user is 10 minutes late for viewing 20/20. Fortunately, the system, based on the three description schemes, has already been recording 20/20 since 10 PM. Now the user can start watching the recorded portion of 20/20 as the recording of 20/20 proceeds. The user will be done viewing 20/20 at 11:10 PM.

20 The average consumer has an ever increasing number of multimedia devices, such as a home audio system, a car stereo, several home television sets, web browsers, etc. The user currently has to customize each of the devices for optimal viewing and/or listening preferences. By storing the user preferences on a removable storage device, such as a smart card, the user may insert the card including the user preferences  
25 into such media devices for automatic customization. This results in the desired programs being automatically recorded on the VCR, and setting of the radio stations for the car

5 stereo and home audio system. In this manner the user only has to specify his preferences at most once, on a single device and subsequently, the descriptors are automatically uploaded into devices by the removable storage device. The user description scheme may also be loaded into other devices using a wired or wireless network connection, e.g. that of a home network. Alternatively, the system can store the user history and create entries in  
10 the user description scheme based on the's audio and video viewing habits. In this manner, the user would never need to program the viewing information to obtain desired information. In a sense, the user descriptor scheme enables modeling of the user by providing a central storage for the user's listening, viewing, browsing preferences, and user's behavior. This enables devices to be quickly personalized, and enables other  
15 components, such as intelligent agents, to communicate on the basis of a standardized description format, and to make smart inferences regarding the user's preferences.

Many different realizations and applications can be readily derived from FIGS. 2 and 3 by appropriately organizing and utilizing their different parts, or by adding peripherals and extensions as needed. In its most general form, FIG. 2 depicts an  
20 audiovisual searching, filtering, browsing, and/or recording appliance that is personalizable. The list of more specific applications/implementations given below is not exhaustive but covers a range.

The user description scheme is a major enabler for personalizable audiovisual appliances. If the structure (syntax and semantics) of the description schemes is  
25 known amongst multiple appliances, the user (user) can carry (or otherwise transfer) the information contained within his user description scheme from one appliance to another,

perhaps via a smart card--where these appliances support smart card interface-- in order to personalize them. Personalization can range from device settings, such as display contrast and volume control, to settings of television channels, radio stations, web stations, web sites, geographic information, and demographic information such as age, zip code etc. Appliances that can be personalized may access content from different sources. They may be connected to the web, terrestrial or cable broadcast, etc., and they may also access multiple or different types of single media such as video, music, etc.

For example, one can personalize the car stereo using a smart card plugged out of the home system and plugged into the car stereo system to be able to tune to favorite stations at certain times. As another example, one can also personalize television viewing, for example, by plugging the smart card into a remote control that in turn will autonomously command the television receiving system to present the user information about current and future programs that fits the user's preferences. Different members of the household can instantly personalize the viewing experience by inserting their own smart card into the family remote. In the absence of such a remote, this same type of personalization can be achieved by plugging in the smart card directly to the television system. The remote may likewise control audio systems. In another implementation, the television receiving system holds user description schemes for multiple users (users) in local storage and identify different users (or group of users) by using an appropriate input interface. For example an interface using user-voice identification technology. It is noted that in a networked system the user description scheme may be transported over the network.

5                   The user description scheme is generated by direct user input, and by using  
a software that watches the user to determine his/her usage pattern and usage history. User  
description scheme can be updated in a dynamic fashion by the user or automatically. A  
well defined and structured description scheme design allows different devices to  
interoperate with each other. A modular design also provides portability.

10                   The description scheme adds new functionality to those of the current VCR.  
An advanced VCR system can learn from the user via direct input of preferences, or by  
watching the usage pattern and history of the user. The user description scheme holds  
user's preferences users and usage history. An intelligent agent can then consult with the  
user description scheme and obtain information that it needs for acting on behalf of the  
15   user. Through the intelligent agent, the system acts on behalf of the user to discover  
programs that fit the taste of the user, alert the user about such programs, and/or record  
them autonomously. An agent can also manage the storage in the system according to the  
user description scheme, i.e., prioritizing the deletion of programs (or alerting the user for  
transfer to a removable media), or determining their compression factor (which directly  
20   impacts their visual quality) according to user's preferences and history.

                  The program description scheme and the system description scheme work in  
collaboration with the user description scheme in achieving some tasks. In addition, the  
program description scheme and system description scheme in an advanced VCR or other  
system will enable the user to browse, search, and filter audiovisual programs. Browsing  
25   in the system offers capabilities that are well beyond fast forwarding and rewinding. For  
instance, the user can view a thumbnail view of different categories of programs stored in

5 the system. The user then may choose frame view, shot view, key frame view, or highlight view, depending on their availability and user's preference. These views can be readily invoked using the relevant information in the program description scheme, especially in program views. The user at any time can start viewing the program either in parts, or in its entirety.

10 In this application, the program description scheme may be readily available from many services such as: (i) from broadcast (carried by EPG defined as a part of ATSC-PSIP (ATSC-Program Service Integration Protocol) in USA or DVB-SI (Digital Video Broadcast-Service Information) in Europe); (ii) from specialized data services (in addition to PSIP/DVB-SI); (iii) from specialized web sites; (iv) from the media storage unit  
15 containing the audiovisual content (e.g., DVD); (v) from advanced cameras (discussed later), and/or may be generated (i.e., for programs that are being stored) by the analysis module 42 or by user input 48.

Contents of digital still and video cameras can be stored and managed by a system that implements the description schemes, e.g., a system as shown in FIG. 2.

20 Advanced cameras can store a program description scheme, for instance, in addition to the audiovisual content itself. The program description scheme can be generated either in part or in its entirety on the camera itself via an appropriate user input interface (e.g., speech, visual menu drive, etc.). Users can input to the camera the program description scheme information, especially those high-level (or semantic) information that may otherwise be  
25 difficult to automatically extract by the system. Some camera settings and parameters (e.g., date and time), as well as quantities computed in the camera (e.g., color histogram to be



5 included in the color profile), can also be used in generating the program description scheme. Once the camera is connected, the system can browse the camera content, or transfer the camera content and its description scheme to the local storage for future use. It is also possible to update or add information to the description scheme generated in the camera.

10 The IEEE 1394 and Havi standard specifications enable this type of “audiovisual content” centric communication among devices. The description scheme API's can be used in the context of Havi to browse and/or search the contents of a camera or a DVD which also contain a description scheme associated with their content, i.e., doing more than merely invoking the PLAY API to play back and linearly view the media.

15 The description schemes may be used in archiving audiovisual programs in a database. The search engine uses the information contained in the program description scheme to retrieve programs on the basis of their content. The program description scheme can also  
20 be used in navigating through the contents of the database or the query results. The user description scheme can be used in prioritizing the results of the user query during presentation. It is possible of course to make the program description scheme more comprehensive depending on the nature of the particular application.

The description scheme fulfills the user's desire to have applications that  
25 pay attention and are responsive to their viewing and usage habits, preferences, and personal demographics. The proposed user description scheme directly addresses this

5        desire in its selection of fields and interrelationship to other description schemes. Because  
the description schemes are modular in nature, the user can port his user description  
scheme from one device to another in order to “personalize” the device.

             The proposed description schemes can be incorporated into current products  
similar to those from TiVo and Replay TV in order to extend their entertainment  
10        informational value. In particular, the description scheme will enable audiovisual  
browsing and searching of programs and enable filtering within a particular program by  
supporting multiple program views such as the highlight view. In addition, the description  
scheme will handle programs coming from sources other than television broadcasts for  
which TiVo and Replay TV are not designed to handle. In addition, by standardization of  
15        TiVo and Replay TV type of devices, other products may be interconnected to such devices  
to extend their capabilities, such as devices supporting an MPEG 7 description. MPEG-7  
is the Moving Pictures Experts Group - 7, acting to standardize descriptions and  
description schemes for audiovisual information. The device may also be extended to be  
personalized by multiple users, as desired.

20                Because the description scheme is defined, the intelligent software agents  
can communicate among themselves to make intelligent inferences regarding the user’s  
preferences. In addition, the development and upgrade of intelligent software agents for  
browsing and filtering applications can be simplified based on the standardized user  
description scheme.

25                The description scheme is multi-modal in the following sense that it holds  
both high level (semantic) and low level features and/or descriptors. For example, the high

5 and low level descriptors are actor name and motion model parameters, respectively. High  
level descriptors are easily readable by humans while low level descriptors are more easily  
read by machines and less understandable by humans. The program description scheme  
can be readily harmonized with existing EPG, PSIP, and DVB-SI information facilitating  
search and filtering of broadcast programs. Existing services can be extended in the future  
10 by incorporating additional information using the compliant description scheme.

For example, one case may include audiovisual programs that are  
prerecorded on a media such as a digital video disc where the digital video disc also  
contains a description scheme that has the same syntax and semantics of the description  
scheme that the FSB module uses. If the FSB module uses a different description scheme,  
15 a transcoder (converter) of the description scheme may be employed. The user may want  
to browse and view the content of the digital video disc. In this case, the user may not need  
to invoke the analysis module to author a program description. However, the user may  
want to invoke his or her user description scheme in filtering, searching and browsing the  
digital video disc content. Other sources of program information may likewise be used in  
20 the same manner.

It is to be understood that any of the techniques described herein with  
relation to video are equally applicable to images (such as still image or a frame of a video)  
and audio (such as radio).

An example of an audiovisual interface is shown in FIGS. 4-12 which is  
25 suitable for the preferred audiovisual description scheme. Referring to FIG. 4, by selecting  
the thumbnail function as a function of category provides a display with a set of categories

on the left hand side. Selecting a particular category, such as news, provides a set of thumbnail views of different programs that are currently available for viewing. In addition, the different programs may also include programs that will be available at a different time for viewing. The thumbnail views are short video segments that provide an indication of the content of the respective actual program that it corresponds with. Referring to FIG. 5, a thumbnail view of available programs in terms of channels may be displayed, if desired. Referring to FIG. 6, a text view of available programs in terms of channels may be displayed, if desired. Referring to FIG. 7, a frame view of particular programs may be displayed, if desired. A representative frame is displayed in the center of the display with a set of representative frames of different programs in the left hand column. The frequency of the number of frames may be selected, as desired. Also a set of frames are displayed on the lower portion of the display representative of different frames during the particular selected program. Referring to FIG. 8, a shot view of particular programs may be displayed, as desired. A representative frame of a shot is displayed in the center of the display with a set of representative frames of different programs in the left hand column. Also a set of shots are displayed on the lower portion of the display representative of different shots (segments of a program, typically sequential in nature) during the particular selected program. Referring to FIG. 9, a key frame view of particular programs may be displayed, as desired. A representative frame is displayed in the center of the display with a set of representative frames of different programs in the left hand column. Also a set of key frame views are displayed on the lower portion of the display representative of different key frame portions during the particular selected program. The number of key

5 frames in each key frame view can be adjusted by selecting the level. Referring to FIG. 10, a highlight view may likewise be displayed, as desired. Referring to FIG. 11, an event view may likewise be displayed, as desired. Referring to FIG. 12, a character/object view may likewise be displayed, as desired.

An example of the description schemes is shown below in XML. The  
10 description scheme may be implemented in any language and include any of the included descriptions (or more), as desired.

The proposed program description scheme includes three major sections for describing a video program. The first section identifies the described program. The second section defines a number of views which may be useful in browsing applications. The third  
15 section defines a number of profiles which may be useful in filtering and search applications. Therefore, the overall structure of the proposed description scheme is as follows:

```
<?XML version="1.0">
<!DOCTYPE MPEG-7 SYSTEM "mpeg-7.dtd">
20<ProgramIdentity>
    <ProgramID> ... </ProgramID>
    <ProgramName> ... </ProgramName>
    <SourceLocation> ... </SourceLocation>
</ProgramIdentity>
25<ProgramViews>
    <ThumbnailView> ... </ThumbnailView>
    <SlideView> ... </SlideView>
    <FrameView> ... </FrameView>
    <ShotView> ... </ShotView>
```

```

5  <KeyFrameView> ... </KeyFrameView>
    <HighlightView> ... </HighlightView>
    <EventView> ... </EventView>
    <CloseUpView> ... </CloseUpView>
    <AlternateView> ... </AlternateView>
10</ProgramViews>
    <ProgramProfiles>
        <GeneralProfile> ... </GeneralProfile>
        <CategoryProfile> ... </CategoryProfile>
        <DateTimeProfile> ... </DateTimeProfile>
15  <KeywordProfile> ... </KeywordProfile>
        <TriggerProfile> ... </TriggerProfile>
        <StillProfile> ... </StillProfile>
        <EventProfile> ... </EventProfile>
        <CharacterProfile> ... </CharacterProfile>
20  <ObjectProfile> ... </ObjectProfile>
        <ColorProfile> ... </ColorProfile>
        <TextureProfile> ... </TextureProfile>
        <ShapeProfile> ... </ShapeProfile>
        <MotionProfile> ... </MotionProfile>
25</ProgramProfiles>

```

## Program Identity

- **Program ID**

```

30  <ProgramID> program-id </ProgramID>

```

The descriptor <ProgramID> contains a number or a string to identify a program.

5      • **Program name**

```
<ProgramName> program-name </ProgramName>
```

The descriptor <ProgramName> specifies the name of a program.

10      • **Source location**

```
<SourceLocation> source-url </SourceLocation>
```

The descriptor <SourceLocation> specifies the location of a program in URL format.

15                      **Program Views**

            • **Thumbnail view**

```
<ThumbnailView>
```

```
    <Image> thumbnail-image </Image>
```

20      </ThumbnailView>

The descriptor <ThumbnailView> specifies an image as the thumbnail representation of a program.

5      • **Slide view**

```
<SlideView> frame-id ... </SlideView>
```

10      The descriptor `<SlideView>` specifies a number of frames in a program which may be viewed as snapshots or in a slide show manner.

• **Frame view**

```
<FrameView> start-frame-id end-frame-id </FrameView>
```

15      The descriptor `<FrameView>` specifies the start and end frames of a program. This is the most basic view of a program and any program has a frame view.

• **Shot view**

```
<ShotView>
```

20      `<Shot id=""> start-frame-id end-frame-id display-frame-id </Shot>`

```
    <Shot id=""> start-frame-id end-frame-id display-frame-id </Shot>
```

```
    ...
```

```
</ShotView>
```



5       The descriptor <ShotView> specifies a number of shots in a program. The <Shot> descriptor defines the start and end frames of a shot. It may also specify a frame to represent the shot.

- **Key-frame view**

```
10  <KeyFrameView>
    <KeyFrames level="">
        <Clip id=""> start-frame-id end-frame-id display-frame-id </Clip>
        <Clip id=""> start-frame-id end-frame-id display-frame-id </Clip>
        ...
15  </KeyFrames>
    <KeyFrames level="">
        <Clip id=""> start-frame-id end-frame-id display-frame-id </Clip>
        <Clip id=""> start-frame-id end-frame-id display-frame-id </Clip>
        ...
20  </KeyFrames>
    ...
</KeyFrameView>
```

25       The descriptor <KeyFrameView> specifies key frames in a program. The key frames may be organized in a hierarchical manner and the hierarchy is captured by the descriptor <KeyFrames> with a level attribute. The clips which are associated with each key frame are defined by the descriptor <Clip>. Here the display frame in each clip is the corresponding key frame.

## 5 • Highlight view

```
<HighlightView>
  <Highlight length="">
    <Clip id=""> start-frame-id end-frame-id display-frame-id </Clip>
10    <Clip id=""> start-frame-id end-frame-id display-frame-id </Clip>
    ...
  </Highlight>
  <Highlight length="">
    <Clip id=""> start-frame-id end-frame-id display-frame-id </Clip>
15    <Clip id=""> start-frame-id end-frame-id display-frame-id </Clip>
    ...
  </Highlight>
  ...
</HighlightView>
```

20

The descriptor <HighlightView> specifies clips to form highlights of a program. A program may have different versions of highlights which are tailored into various time length. The clips are grouped into each version of highlight which is specified by the descriptor <Highlight> with a length attribute.

## 25 • Event view

```
<EventView>
  <Events name="">
    <Clip id=""> start-frame-id end-frame-id display-frame-id </Clip>
30    <Clip id=""> start-frame-id end-frame-id display-frame-id </Clip>
    ...
```

```

5      </Events>
      <Events name="">
        <Clip id=""> start-frame-id end-frame-id display-frame-id </Clip>
        <Clip id=""> start-frame-id end-frame-id display-frame-id </Clip>
        ...
10     </Events>
        ...
</EventView>

```

The descriptor <EventView> specifies clips which are related to certain events in a  
 15 program. The clips are grouped into the corresponding events which are specified by the  
 descriptor <Event> with a name attribute.

- **Close-up view**

```

<CloseUpView>
20   <Target name="">
        <Clip id=""> start-frame-id end-frame-id display-frame-id </Clip>
        <Clip id=""> start-frame-id end-frame-id display-frame-id </Clip>
        ...
    </Target>
25   <Target name="">
        <Clip id=""> start-frame-id end-frame-id display-frame-id </Clip>
        <Clip id=""> start-frame-id end-frame-id display-frame-id </Clip>
        ...
    </Target>
30   ...
</CloseUpView>

```

5           The descriptor <CloseUpView> specifies clips which may be zoomed in to certain targets in a program. The clips are grouped into the corresponding targets which are specified by the descriptor <Target> with a name attribute.

- **Alternate view**

10   <AlternateView>  
      <AlternateSource id=""> source-url </AlternateSource>  
      <AlternateSource id=""> source-url </AlternateSource>  
      ...  
  </AlternateView>

15           The descriptor <AlternateView> specifies sources which may be shown as alternate views of a program. Each alternate view is specified by the descriptor <AlternateSource> with an id attribute. The locate of the source may be specified in URL format.

## **Program Profiles**

20       • **General profile**

      <GeneralProfile>  
          <Title> title-text </Title>  
          <Abstract> abstract-text </Abstract>  
25       <Audio> voice-annotation </Audio>  
          <Www> web-page-url </Www>

5       <ClosedCaption> yes/no </ClosedCaption>

          <Language> language-name </Language>

          <Rating> rating </Rating>

          <Length> time </Length>

          <Authors> author-name ... </Authors>

10       <Producers> producer-name ... </Producers>

          <Directors> director-name ... </Directors>

          <Actors> actor-name ... </Actors>

          ...

      </GeneralProfile>

15

The descriptor <GeneralProfile> describes the general aspects of a program.

- **Category profile**

<CategoryProfile> category-name ... </CategoryProfile>

20

The descriptor <CategoryProfile> specifies the categories under which a program may be classified.

- **Date-time profile**

25   <DateTimeProfile>

      <ProductionDate> date </ProductionDate>

      <ReleaseDate> date </ReleaseDate>

      <RecordingDate> date </RecordingDate>

      <RecordingTime> time </RecordingTime>

5       ...

`</DateTimeProfile>`

The descriptor `<DateTimeProfile>` specifies various date and time information of a program.

10       • **Keyword profile**

`<KeywordProfile> keyword ... </KeywordProfile>`

15       The descriptor `<KeywordProfile>` specifies a number of keywords which may be used to filter or search a program.

• **Trigger profile**

`<TriggerProfile> trigger-frame-id ... </TriggerProfile>`

20       The descriptor `<TriggerProfile>` specifies a number of frames in a program which may be used to trigger certain actions while the playback of the program.

• **Still profile**

`<StillProfile>`

```

5      <Still id="">
      <HotRegion id="">
          <Location> x1 y1 x2 y2 </Location>
          <Text> text-annotation </Text>
          <Audio> voice-annotation </Audio>
10      <Www> web-page-url </Www>
      </HotRegion>
      <HotRegion id="">
          <Location> x1 y1 x2 y2 </Location>
          <Text> text-annotation </Text>
15      <Audio> voice-annotation </Audio>
          <Www> web-page-url </Www>
      </HotRegion>
      ...
</Still>
20 <Still id="">
      <HotRegion id="">
          <Location> x1 y1 x2 y2 </Location>
          <Text> text-annotation </Text>
          <Audio> voice-annotation </Audio>
25      <Www> web-page-url </Www>
      </HotRegion>
      <HotRegion id="">
          <Location> x1 y1 x2 y2 </Location>
          <Text> text-annotation </Text>
30      <Audio> voice-annotation </Audio>
          <Www> web-page-url </Www>
      </HotRegion>
      ...
</Still>
35 ...
</StillProfile>

```

The descriptor `<StillProfile>` specifies hot regions or regions of interest within a frame. The frame is specified by the descriptor `<Still>` with an `id` attribute which corresponds to the frame-id. Within a frame, each hot region is specified by the descriptor `<HotRegion>` with an `id` attribute.

## 10 • Event profile

```

<EventProfile>
  <EventList> event-name ... </EventList>
  <Event name="">
15    <Www> web-page-url </Www>
    <Occurrence id="">
      <Duration> start-frame-id end-frame-id </Duration>
      <Text> text-annotation </Text>
      <Audio> voice-annotation </Audio>
20    </Occurrence>
    <Occurrence id="">
      <Duration> start-frame-id end-frame-id </Duration>
      <Text> text-annotation </Text>
      <Audio> voice-annotation </Audio>
25    </Occurrence>
    ...
  </Event>
  <Event name="">
    <Www> web-page-url </Www>
30    <Occurrence id="">
      <Duration> start-frame-id end-frame-id </Duration>

```



```

5          <Text> text-annotation </Text>
          <Audio> voice-annotation </Audio>
        </Occurrence>
        <Occurrence id="">
          <Duration> start-frame-id end-frame-id </Duration>
10        <Text> text-annotation </Text>
          <Audio> voice-annotation </Audio>
        </Occurrence>
        ...
      </Event>
15    ...
  </EventProfile>

```

The descriptor <EventProfile> specifies the detailed information for certain events in a program. Each event is specified by the descriptor <Event> with a name attribute. Each occurrence of an event is specified by the descriptor <Occurrence> with an id attribute which may be matched with a clip id under <EventView>.

- **Character profile**

```

<CharacterProfile>
25  <CharacterList> character-name ... </CharacterList>
    <Character name="">
      <ActorName> actor-name </ActorName>
      <Gender> male </Gender>
      <Age> age </Age>
30  <Www> web-page-url </Www>
    <Occurrence id="">

```

5           <Duration> start-frame-id end-frame-id </Duration>

            <Location> frame:[x1 y1 x2 y2] ... </Location>

            <Motion>  $v_x$   $v_y$   $v_z$   $v_\alpha$   $v_\beta$   $v_\gamma$  </Motion>

            <Text> text-annotation </Text>

            <Audio> voice-annotation </Audio>

10          </Occurrence>

            <Occurrence id="">

                <Duration> start-frame-id end-frame-id </Duration>

                <Location> frame:[x1 y1 x2 y2] ... </Location>

                <Motion>  $v_x$   $v_y$   $v_z$   $v_\alpha$   $v_\beta$   $v_\gamma$  </Motion>

15          <Text> text-annotation </Text>

            <Audio> voice-annotation </Audio>

            </Occurrence>

            ...

            </Character>

20          <Character name="">

            <ActorName> actor-name </ActorName>

            <Gender> male </Gender>

            <Age> age </Age>

            <Www> web-page-url </Www>

25          <Occurrence id="">

            <Duration> start-frame-id end-frame-id </Duration>

            <Location> frame:[x1 y1 x2 y2] ... </Location>

            <Motion>  $v_x$   $v_y$   $v_z$   $v_\alpha$   $v_\beta$   $v_\gamma$  </Motion>

            <Text> text-annotation </Text>

30          <Audio> voice-annotation </Audio>

            </Occurrence>

            <Occurrence id="">

                <Duration> start-frame-id end-frame-id </Duration>

                <Location> frame:[x1 y1 x2 y2] ... </Location>

35          <Motion>  $v_x$   $v_y$   $v_z$   $v_\alpha$   $v_\beta$   $v_\gamma$  </Motion>

            <Text> text-annotation </Text>

```

5          <Audio> voice-annotation </Audio>

        </Occurrence>

        ...

    </Character>

    ...
10 </CharacterProfile>

```

The descriptor <CharacterProfile> specifies the detailed information for certain characters in a program. Each character is specified by the descriptor <Character> with a name attribute. Each occurrence of a character is specified by the descriptor <Occurrence> with an id attribute which may be matched with a clip id under <CloseUpView>.

## - Object profile

```

<ObjectProfile>
    <ObjectList> object-name ... </ObjectList>
20 <Object name="">
    <Www> web-page-url </Www>
    <Occurrence id="">
        <Duration> start-frame-id end-frame-id </Duration>
        <Location> frame:[x1 y1 x2 y2] ... </Location>
25 <Motion> vx vy vz vα vβ vγ </Motion>
        <Text> text-annotation </Text>
        <Audio> voice-annotation </Audio>
    </Occurrence>
    <Occurrence id="">
30 <Duration> start-frame-id end-frame-id </Duration>
        <Location> frame:[x1 y1 x2 y2] ... </Location>

```

```

5      <Motion> vx vy vz vα vβ vγ </Motion>

      <Text> text-annotation </Text>

      <Audio> voice-annotation </Audio>

    </Occurrence>

    ...

10  </Object>

    <Object name="">

      <Www> web-page-url </Www>

      <Occurrence id="">

        <Duration> start-frame-id end-frame-id </Duration>

15      <Location> frame:[x1 y1 x2 y2] ... </Location>

        <Motion> vx vy vz vα vβ vγ </Motion>

        <Text> text-annotation </Text>

        <Audio> voice-annotation </Audio>

      </Occurrence>

20      <Occurrence id="">

        <Duration> start-frame-id end-frame-id </Duration>

        <Location> frame:[x1 y1 x2 y2] ... </Location>

        <Motion> vx vy vz vα vβ vγ </Motion>

        <Text> text-annotation </Text>

25      <Audio> voice-annotation </Audio>

      </Occurrence>

      ...

    </Object>

    ...

30 </ObjectProfile>

```

The descriptor <ObjectProfile> specifies the detailed information for certain objects in a program. Each object is specified by the descriptor <Object> with a name attribute. Each occurrence of a object is specified by the descriptor <Occurrence> with an id attribute

35 which may be matched with a clip id under <CloseUpView>.

## 5 • Color profile

```
<ColorProfile>  
    ...  
</ColorProfile>
```

10

The descriptor <ColorProfile> specifies the detailed color information of a program. All MPEG-7 color descriptors may be placed under here.

## • Texture profile

```
15 <TextureProfile>  
    ...  
</TextureProfile>
```

The descriptor <TextureProfile> specifies the detailed texture information of a program.

20 All MPEG-7 texture descriptors may be placed under here.

## • Shape profile

```
<ShapeProfile>  
    ...  
25 </ShapeProfile>
```

5           The descriptor <ShapeProfile> specifies the detailed shape information of a program. All  
MPEG-7 shape descriptors may be placed under here.

- **Motion profile**

```
10           <MotionProfile>  
            ...  
          </MotionProfile>
```

The descriptor <MotionProfile> specifies the detailed motion information of a program.  
All MPEG-7 motion descriptors may be placed under here.

## 15           **User Description Scheme**

The proposed user description scheme includes three major sections for describing a user. The first section identifies the described user. The second section records a number of settings which may be preferred by the user. The third section records some statistics which may reflect certain usage patterns of the user. Therefore, the overall structure of the proposed description  
20       scheme is as follows:

```
<?XML version="1.0">  
<!DOCTYPE MPEG-7 SYSTEM "mpeg-7.dtd">  
<UserIdentity>  
    <UserID> ... </UserID>
```

```

5  <UserName> ... </UserName>

   </UserIdentity>

   <UserPreferences>

       <BrowsingPreferences> ... </BrowsingPreferences>

       <FilteringPreferences> ... </FilteringPreferences>
10  <SearchPreferences> ... </SearchPreferences>

       <DevicePreferences> ... </DevicePreferences>

   </UserPreferences>

   <UserHistory>

       <BrowsingHistory> ... </BrowsingHistory>
15  <FilteringHistory> ... </FilteringHistory>

       <SearchHistory> ... </SearchHistory>

       <DeviceHistory> ... </DeviceHistory>

   </UserHistory>

   <UserDemographics>
20  <Age> ... </Age>

       <Gender> ... </Gender>

       <ZIP> ... </ZIP>

   </UserDemographics>

```

## 25            **User Identity**

- **User ID**

```
<UserID> user-id </UserID>
```

30            The descriptor <UserID> contains a number or a string to identify a user.

## 5 • User name

```
<UserName> user-name </UserName>
```

The descriptor `<UserName>` specifies the name of a user.

## 10 User Preferences

### • Browsing preferences

```
<BrowsingPreferences>
  <Views>
    <ViewCategory id=""> view-id ... </ViewCategory>
    <ViewCategory id=""> view-id ... </ViewCategory>
    ...
  </Views>
  <FrameFrequency> frequency ...<FrameFrequency>
  <ShotFrequency> frequency ...<ShotFrequency>
  <KeyFrameLevel> level-id ...<KeyFrameLevel>
  <HighlightLength> length ...<HighlightLength>
  ...
</BrowsingPreferences>
```

25

The descriptor `<BrowsingPreferences>` specifies the browsing preferences of a user. The user's preferred views are specified by the descriptor `<Views>`. For each category, the preferred views are specified by the descriptor `<ViewCategory>` with an `id` attribute which corresponds to the category id. The descriptor `<FrameFrequency>` specifies at what



5 interval the frames should be displayed on a browsing slider under the frame view. The  
descriptor <ShotFrequency> specifies at what interval the shots should be displayed on a  
browsing slider under the shot view. The descriptor <KeyFrameLevel> specifies at what  
level the key frames should be displayed on a browsing slider under the key frame view.  
The descriptor <HighlightLength> specifies which version of the highlight should be  
10 shown under the highlight view.

- **Filtering preferences**

```
<FilteringPreferences>
    <Categories> category-name ... </Categories>
    <Channels> channel-number ... </Channels>
    <Ratings> rating-id ... </Ratings>
    <Shows> show-name ... </Shows>
    <Authors> author-name ... </Authors>
    <Producers> producer-name ... </Producers>
    <Directors> director-name ... </Directors>
    <Actors> actor-name ... </Actors>
    <Keywords> keyword ... </Keywords>
    <Titles> title-text ... </Titles>
    ...
25 </FilteringPreferences>
```

The descriptor <FilteringPreferences> specifies the filtering related preferences of a user.

## 5 • Search preferences

```
<SearchPreferences>
    <Categories> category-name ... </Categories>
    <Channels> channel-number ... </Channels>
10    <Ratings> rating-id ... </Ratings>
    <Shows> show-name ... </Shows>
    <Authors> author-name ... </Authors>
    <Producers> producer-name ... </Producers>
    <Directors> director-name ... </Directors>
15    <Actors> actor-name ... </Actors>
    <Keywords> keyword ... </Keywords>
    <Titles> title-text ... </Titles>
    ...
20 </SearchPreferences>
```

The descriptor `<SearchPreferences>` specifies the search related preferences of a user.

## • Device preferences

```
<DevicePreferences>
25    <Brightness> brightness-value </Brightness>
    <Contrast> contrast-value </Contrast>
    <Volume> volume-value </Volume>
</DevicePreferences>
```

30 The descriptor `<DevicePreferences>` specifies the device preferences of a user.

- **Browsing history**

```

<BrowsingHistory>
  <Views>
10    <ViewCategory id=""> view-id ... </ViewCategory>
      <ViewCategory id=""> view-id ... </ViewCategory>
      ...
  </Views>
  <FrameFrequency> frequency ...<FrameFrequency>
15  <ShotFrequency> frequency ...<ShotFrequency>
      <KeyFrameLevel> level-id ...<KeyFrameLevel>
      <HighlightLength> length ...<HighlightLength>
      ...
20 </BrowsingHistory>

```

The descriptor <BrowsingHistory> captures the history of a user's browsing related activities.

- **Filtering history**

```

25 <FilteringHistory>
    <Categories> category-name ... </Categories>
    <Channels> channel-number ... </Channels>
    <Ratings> rating-id ... </Ratings>
    <Shows> show-name ... </Shows>
30 <Authors> author-name ... </Authors>

```

```

5      <Producers> producer-name ... </Producers>

      <Directors> director-name ... </Directors>

      <Actors> actor-name ... </Actors>

      <Keywords> keyword ... </Keywords>

      <Titles> title-text ... </Titles>
10     ...

    </FilteringHistory>

```

The descriptor <FilteringHistory> captures the history of a user's filtering related activities.

## 15 • Search history

```

    <SearchHistory>

      <Categories> category-name ... </Categories>

      <Channels> channel-number ... </Channels>
20    <Ratings> rating-id ... </Ratings>

      <Shows> show-name ... </Shows>

      <Authors> author-name ... </Authors>

      <Producers> producer-name ... </Producers>

      <Directors> director-name ... </Directors>

      <Actors> actor-name ... </Actors>
25    <Keywords> keyword ... </Keywords>

      <Titles> title-text ... </Titles>

      ...

    </SearchHistory>
30

```

The descriptor <SearchHistory> captures the history of a user's search related activities.

## 5 • Device history

```
<DeviceHistory>
  <Brightness> brightness-value ... </Brightness>
  <Contrast> contrast-value ... </Contrast>
10  <Volume> volume-value ... </Volume>
</DeviceHistory>
```

The descriptor `<DeviceHistory>` captures the history of a user's device related activities.

## User demographics

### 15 • Age

```
<Age> age </Age>
```

The descriptor `<Age>` specifies the age of a user.

### 20 • Gender

```
<Gender> ... </Gender>
```

The descriptor `<Gender>` specifies the gender of a user.

## 5 • ZIP code

```
<ZIP> ... </ZIP>
```

The descriptor <ZIP> specifies the ZIP code of where a user lives.

## 10 System Description Scheme

The proposed system description scheme includes four major sections for describing a user.

The first section identifies the described system. The second section keeps a list of all known users. The third section keeps lists of available programs. The fourth section describes the capabilities of the system. Therefore, the overall structure of the proposed description scheme

15 is as follows:

```
<?XML version="1.0">
<!DOCTYPE MPEG-7 SYSTEM "mpeg-7.dtd">
<SystemIdentity>
  <SystemID> ... </SystemID>
20 <SystemName> ... </SystemName>
  <SystemSerialNumber> ... </SystemSerialNumber>
</SystemIdentity>
<SystemUsers>
  <Users> ... </Users>
25</SystemUsers>
<SystemPrograms>
  <Categories> ... </Categories>
  <Channels> ... </Channels>
```

```

5  <Programs> ... </Programs>
   </SystemPrograms>
   <SystemCapabilities>
       <Views> ... </Views>
   </SystemCapabilities>

```

10

## System Identity

- **System ID**

```
<SystemID> system-id </SystemID>
```

15

The descriptor `<SystemID>` contains a number or a string to identify a video system or device.

- **System name**

```
<SystemName> system-name </SystemName>
```

20

The descriptor `<SystemName>` specifies the name of a video system or device.

- **System serial number**

```
<SystemSerialNumber> system-serial-number </SystemSerialNumber>
```

25

5       The descriptor <SystemSerialNumber> specifies the serial number of a video system or device.

## System Users

- **Users**

```
10  <Users>
    <User>
        <UserID> user-id </UserID>
        <UserName> user-name </UserName>
    </User>
15  <User>
        <UserID> user-id </UserID>
        <UserName> user-name </UserName>
    </User>
    ...
20  </Users>
```

The descriptor <SystemUsers> lists a number of users who have registered on a video system or device. Each user is specified by the descriptor <User>. The descriptor <UserID> specifies a number or a string which should match with the number or string specified in <UserID> in one of the user description schemes.

25



## 5 Programs in the System

- **Categories**

```
<Categories>
  <Category>
10    <CategoryID> category-id </CategoryID>
      <CategoryName> category-name </CategoryName>
      <SubCategories> sub-category-id ... </SubCategories>
  </Category>
  <Category>
15    <CategoryID> category-id </CategoryID>
      <CategoryName> category-name </CategoryName>
      <SubCategories> sub-category-id ... </SubCategories>
  </Category>
  ...
20 </Categories>
```

The descriptor <Categories> lists a number of categories which have been registered on a video system or device. Each category is specified by the descriptor <Category>. The major-sub relationship between categories is captured by the descriptor <SubCategories>.

## 25 • Channels

```
<Channels>
  <Channel>
    <ChannelID> channel-id </ChannelID>
```

```

5      <ChannelName> channel-name </ChannelName>

      <SubChannels> sub-channel-id ... </SubChannels>

</Channel>

<Channel>

      <ChannelID> channel-id </ChannelID>

10     <ChannelName> channel-name </ChannelName>

      <SubChannels> sub-channel-id ... </SubChannels>

</Channel>

...

</Channels>

```

The descriptor <Channels> lists a number of channels which have been registered on a video system or device. Each channel is specified by the descriptor <Channel>. The major-sub relationship between channels is captured by the descriptor <SubChannels>.

## • Programs

```

20 <Programs>

    <CategoryPrograms>

        <CategoryID> category-id </CategoryID>

        <Programs> program-id ... </Programs>

25 </CategoryPrograms>

    <CategoryPrograms>

        <CategoryID> category-id </CategoryID>

        <Programs> program-id ... </Programs>

    </CategoryPrograms>

30 ...

    <ChannelPrograms>

        <ChannelID> channel-id </ChannelID>

```

```

5      <Programs> program-id ... </Programs>
      </ChannelPrograms>
      <ChannelPrograms>
          <ChannelID> channel-id </ChannelID>
          <Programs> program-id ... </Programs>
10     </ChannelPrograms>
      ...
      </Programs>

```

The descriptor <Programs> lists programs who are available on a video system or device.

15 The programs are grouped under corresponding categories or channels. Each group of programs are specified by the descriptor <CategoryPrograms> or <ChannelPrograms>. Each program id contained in the descriptor <Programs> should match with the number or string specified in <ProgramID> in one of the program description schemes.

## System Capabilities

### 20 • Views

```

      <Views>
          <View>
              <ViewID> view-id </ViewID>
25      <ViewName> view-name </ViewName>
          </View>
          <View>
              <ViewID> view-id </ViewID>
              <ViewName> view-name </ViewName>

```

5       </view>  
      ...  
      </Views>

The descriptor <Views> lists views which are supported by a video system or device.

10       Each view is specified by the descriptor <View>. The descriptor <ViewName>  
      contains a string which should match with one of the following views used in the  
      program description schemes: ThumbnailView, SlideView, FrameView, ShotView,  
      KeyFrameView, HighlightView, EventView, and CloseUpView.

15               The present inventors came to the realization that the program description  
      scheme may be further modified to provide additional capabilities. Referring to FIG. 13,  
      the modified program description scheme 400 includes four separate types of information,  
      namely, a syntactic structure description scheme 402, a semantic structure description  
      scheme 404, a visualization description scheme 406, and a meta information description  
20       scheme 408. It is to be understood that in any particular system one or more of the  
      description schemes may be included, as desired.

      Referring to FIG. 14, the visualization description scheme 406 enables fast  
      and effective browsing of video program (and audio programs) by allowing access to the  
      necessary data, preferably in a one-step process. The visualization description scheme 406  
25       provides for several different presentations of the video content (or audio), such as for  
      example, a thumbnail view description scheme 410, a key frame view description scheme  
      412, a highlight view description scheme 414, an event view description scheme 416, a

5 close-up view description scheme 418, and an alternative view description scheme 420. Other presentation techniques and description schemes may be added, as desired. The thumbnail view description scheme 410 preferably includes an image 422 or reference to an image representative of the video content and a time reference 424 to the video. The key frame view description scheme 412 preferably includes a level indicator 426 and a time  
10 reference 428. The level indicator 426 accommodates the presentation of a different number of key frames for the same video portion depending on the user's preference. The highlight view description scheme 414 includes a length indicator 430 and a time reference 432. The length indicator 430 accommodates the presentation of a different highlight duration of a video depending on the user's preference. The event view description scheme  
15 416 preferably includes an event indicator 434 for the selection of the desired event and a time reference 436. The close-up view description scheme 418 preferably includes a target indicator 438 and a time reference 440. The alternate view description scheme preferably includes a source indicator 442. To increase performance of the system it is preferred to specify the data which is needed to render such views in a centralized and straightforward  
20 manner. By doing so, it is then feasible to access the data in a simple one-step process without complex parsing of the video.

Referring to FIG. 15, the meta information description scheme 408 generally includes various descriptors which carry general information about a video (or audio) program such as the title, category, keywords, etc. Additional descriptors, such as  
25 those previously described, may be included, as desired.

5 Referring again to FIG. 13, the syntactic structure description scheme 402 specifies the physical structure of a video program (or audio), e.g., a table of contents. The physical features, may include for example, color, texture, motion, etc. The syntactic structure description scheme 402 preferably includes three modules, namely a segment description scheme 450, a region description scheme 452, and a segment/region relation graph description scheme 454. The segment description scheme 450 may be used to define relationships between different portions of the video consisting of multiple frames of the video. A segment description scheme 450 may contain another segment description scheme 450 and/or shot description scheme to form a segment tree. Such a segment tree may be used to define a temporal structure of a video program. Multiple segment trees may be created and thereby create multiple table of contents. For example, a video program may be segmented into story units, scenes, and shots, from which the segment description scheme 450 may contain such information as a table of contents. The shot description scheme may contain a number of key frame description schemes, a mosaic description scheme(s), a camera motion description scheme(s), etc. The key frame description scheme may contain a still image description scheme which may in turn contains color and texture descriptors. It is noted that various low level descriptors may be included in the still image description scheme under the segment description scheme. Also, the visual descriptors may be included in the region description scheme which is not necessarily under a still image description scheme. On example of a segment description scheme 450 is shown in FIG. 16.

5 Referring to FIG. 17, the region description scheme 452 defines the interrelationships between groups of pixels of the same and/or different frames of the video. The region description scheme 452 may also contain geometrical features, color, texture features, motion features, etc.

10 Referring to FIG. 18, the segment/region relation graph description scheme 454 defines the interrelationships between a plurality of regions (or region description schemes), a plurality of segments (or segment description schemes), and/or a plurality of regions (or description schemes) and segments (or description schemes).

15 Referring again to FIG. 13, the semantic structure description scheme 404 is used to specify semantic features of a video program (or audio), e.g. semantic events. In a similar manner to the syntactic structure description scheme, the semantic structure description scheme 404 preferably includes three modules, namely an event description scheme 480, an object description scheme 482, and an event/objection relation graph description scheme 484. The event description scheme 480 may be used to form relationships between different events of the video normally consisting of multiple frames  
20 of the video. An event description scheme 480 may contain another event description scheme 480 to form a segment tree. Such an event segment tree may be used to define a semantic index table for a video program. Multiple event trees may be created and thereby creating multiple index tables. For example, a video program may include multiple events, such as a basketball dunk, a fast break, and a free throw, and the event description scheme  
25 may contain such information as an index table. The event description scheme may also contain references which link the event to the corresponding segments and/or regions

5 specified in the syntactic structure description scheme. On example of an event description scheme is shown in FIG. 19.

Referring to FIG. 20, the object description scheme 482 defines the interrelationships between groups of pixels of the same and/or different frames of the video representative of objects. The object description scheme 482 may contain another object  
10 description scheme and thereby form an object tree. Such an object tree may be used to define an object index table for a video program. The object description scheme may also contain references which link the object to the corresponding segments and/or regions specified in the syntactic structure description scheme.

Referring to FIG. 21, the event/object relation graph description scheme 484  
15 defines the interrelationships between a plurality of events (or event description schemes), a plurality of objects (or object description schemes), and/or a plurality of events (or description schemes) and objects (or description schemes).

The terms and expressions that have been employed in the foregoing specification are used as terms of description and not of limitation, and there is no  
20 intention, in the use of such terms and expressions, of excluding equivalents of the features shown and described or portions thereof, it being recognized that the scope of the invention is defined and limited only by the claims that follow.